

# Physics-Informed Echo State Networks for Chaotic Systems Forecasting<sup>\*</sup>

N.A.K. Doan<sup>1,2</sup>, W. Polifke<sup>1</sup>, and L. Magri<sup>2,3</sup>

<sup>1</sup> Department of Mechanical Engineering, Technical University of Munich, Germany

<sup>2</sup> Institute for Advanced Study, Technical University of Munich, Germany

<sup>3</sup> Department of Engineering, University of Cambridge, United Kingdom

**Abstract.** We propose a physics-informed Echo State Network (ESN) to predict the evolution of chaotic systems. Compared to conventional ESNs, the physics-informed ESNs are trained to solve supervised learning tasks while ensuring that their predictions do not violate physical laws. This is achieved by introducing an additional loss function during the training of the ESNs, which penalizes non-physical predictions without the need of any additional training data. This approach is demonstrated on a chaotic Lorenz system, where the physics-informed ESNs improve the predictability horizon by about two Lyapunov times as compared to conventional ESNs. The proposed framework shows the potential of using machine learning combined with prior physical knowledge to improve the time-accurate prediction of chaotic dynamical systems.

**Keywords:** Echo State Networks · Physics-Informed Neural Networks · Chaotic Dynamical Systems.

## 1 Introduction

Over the past few years, there has been a rapid increase in the development of machine learning techniques, which have been applied with success to various disciplines, from image or speech recognition [2, 5] to playing Go [14]. However, the application of such methods to the study and forecasting of physical systems has only been recently explored including some applications in the field of fluid dynamics [4, 6, 13, 15]. One of the major challenges for using machine learning algorithms for the study of complex physical systems is the prohibitive cost of data acquisition and generation for training [1, 12]. However, in complex physical systems, there exists a large amount of prior knowledge, which can be exploited to improve existing machine learning approaches. These hybrid approaches, called *physics-informed machine learning*, have been applied with some success to flow-structure interaction problems [12], turbulence modelling [6] and the solution of partial differential equations (PDEs) [12].

---

<sup>\*</sup> The authors acknowledge the support of the Technical University of Munich - Institute for Advanced Study, funded by the German Excellence Initiative and the European Union Seventh Framework Programme under grant agreement no. 291763. L.M. also acknowledges the Royal Academy of Engineering Research Fellowship Scheme.

In this study, we propose an approach to combine physical knowledge with a machine learning algorithm to time-accurately forecast the evolution of a chaotic dynamical system. The machine learning tools we use are based on reservoir computing [9], in particular, Echo State Networks (ESNs). ESNs have shown to predict nonlinear and chaotic dynamics more accurately and for a longer time horizon than other deep learning algorithms [9]. ESNs have also recently been used to predict the evolution of spatiotemporal chaotic systems [10, 11]. In the present study, ESNs are augmented by physical constraints to accurately forecast the evolution of a prototypical chaotic system, i.e., the Lorenz system [7].

Sections 2 details the method used for the training and for forecasting the dynamical systems, both with conventional ESNs and the newly proposed physics-informed ESNs. Results are presented in section 3 and final comments are summarized in section 4.

## 2 Methodology

The Echo State Network (ESN) approach presented in [8] is used here. Given a training input signal  $\mathbf{u}(n)$  of dimension  $N_u$  and a desired known target output signal  $\mathbf{y}(n)$  of dimension  $N_y$ , the ESN has to learn a model with output  $\hat{\mathbf{y}}(n)$  matching  $\mathbf{y}(n)$ .  $n = 1, \dots, N_t$  is the discrete time and  $N_t$  is the number of data points in the training dataset covering a discrete time from 0 until time  $T = (N_t - 1)\Delta t$ . In the particular case studied here, where the forecasting of a dynamical system is of interest, the desired output signal is equal to the input signal at the next time step, i.e.,  $\mathbf{y}(n) = \mathbf{u}(n + 1)$ .

The ESN is composed of an artificial high dimensional dynamical system, called a reservoir, whose states at time  $n$  are represented by a vector,  $\mathbf{x}(n)$ , of dimension  $N_x$ , representing the reservoir neuron activations. This reservoir is coupled to the input signal,  $\mathbf{u}$ , via an input-to-reservoir matrix,  $\mathbf{W}_{in}$  of dimension  $N_x \times N_u$ . The output of the reservoir,  $\hat{\mathbf{y}}$ , is deduced from the states via the reservoir-to-output matrix,  $\mathbf{W}_{out}$  of dimension  $N_y \times N_x$ , as a linear combination of the reservoir states:

$$\hat{\mathbf{y}} = \mathbf{W}_{out}\mathbf{x} \quad (1)$$

In this work, a non-leaky reservoir is used, where the state of the reservoir evolves according to:

$$\mathbf{x}(n + 1) = \tanh(\mathbf{W}_{in}\mathbf{u}(n) + \mathbf{W}\mathbf{x}(n)) \quad (2)$$

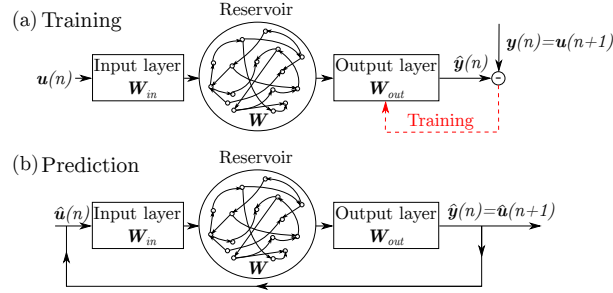
where  $\mathbf{W}$  is the recurrent weight matrix of dimension  $N_x \times N_x$  and the (element-wise)  $\tanh$  function is used as an activation function for the reservoir neurons.

In the conventional ESN approach, illustrated in Fig. 1a, the input and recurrent matrices,  $\mathbf{W}_{in}$  and  $\mathbf{W}$ , are randomly initialized once and are not further trained. These are typically sparse matrices constructed so that the reservoir verifies the Echo State Property [3]. Only the output matrix,  $\mathbf{W}_{out}$ , is then trained to minimize the mean-square-error,  $E_d$ , between the ESN predictions and the data defined as:

$$E_d = \frac{1}{N_y} \sum_{i=1}^{N_y} \frac{1}{N_t} \sum_{n=1}^{N_t} (\hat{y}_i(n) - y_i(n))^2 \quad (3)$$

The subscript  $d$  is used to indicate the error based on the available data.

In the present implementation, following [11],  $\mathbf{W}_{in}$  is generated such that each row of the matrix has only one randomly chosen nonzero element, which is independently taken from a uniform distribution in the interval  $[-\sigma_{in}, \sigma_{in}]$ .  $\mathbf{W}$  is constructed to have an average connectivity  $\langle d \rangle$  and the non-zero elements are taken from a uniform distribution over the interval  $[-1, 1]$ . All the coefficients of  $\mathbf{W}$  are then multiplied by a constant coefficient for the largest absolute eigenvalue of  $\mathbf{W}$  to be equal to a value  $\Lambda$ .



**Fig. 1.** Schematic of the ESN during (a) training and (b) future prediction. The physical constraints are imposed during the training phase (a).

After training, to obtain predictions from the ESN for future time  $t > T$ , the output of the ESN is looped back as the input of the ESN and it then evolves autonomously as represented in Fig. 1b.

## 2.1 Training

As discussed earlier, the training of the ESN consists of the optimization of  $\mathbf{W}_{out}$ . As the outputs of the ESN,  $\hat{\mathbf{y}}$ , are a linear combination of the states,  $\mathbf{x}$ ,  $\mathbf{W}_{out}$  can be obtained by using a simple Ridge regression:

$$\mathbf{W}_{out} = \mathbf{Y} \mathbf{X}^T (\mathbf{X} \mathbf{X}^T + \gamma \mathbf{I})^{-1} \quad (4)$$

where  $\mathbf{Y}$  and  $\mathbf{X}$  are respectively the column-concatenation of the various time instants of the output data,  $\mathbf{y}$ , and associated ESN states  $\mathbf{x}$ .  $\gamma$  is the Tikhonov regularization factor, which helps avoid overfitting. Explicitly, the optimization in Eq. (4) reads:

$$\mathbf{W}_{out} = \underset{\mathbf{W}_{out}}{\operatorname{argmin}} \frac{1}{N_y} \sum_{i=1}^{N_y} \left( \sum_{n=1}^{N_t} (\hat{y}_i(n) - y_i(n))^2 + \gamma \|\mathbf{w}_{out,i}\|^2 \right) \quad (5)$$

where  $\mathbf{w}_{out,i}$  denotes the  $i$ -th row of  $\mathbf{W}_{out}$ . This optimization problem penalizes large values of  $\mathbf{W}_{out}$ , generally improves the feedback stability and avoids overfitting [8].

In this work, following the approach of [12] for artificial deep feedforward neural networks, we propose an alternative approach to training  $\mathbf{W}_{out}$ , which combines the data available with prior physical knowledge of the system under investigation. Let us first assume that the dynamical system is governed by the following nonlinear differential equation:

$$\mathcal{F}(\mathbf{y}) \equiv \partial_t \mathbf{y} + \mathcal{N}(\mathbf{y}) = 0 \quad (6)$$

where  $\mathcal{F}$  is a general non-linear operator,  $\partial_t$  is the time derivative and  $\mathcal{N}$  is a nonlinear differential operator. Eq. (6) represents a formal equation describing the dynamics of a generic nonlinear system. The training phase can be reframed to make use of our knowledge of  $\mathcal{F}$  by minimising the mean squared error,  $E_d$ , and a physical error,  $E_p$ , based on  $\mathcal{F}$ :

$$E_{tot} = E_d + E_p, \text{ where } E_p = \frac{1}{N_y} \sum_{i=1}^{N_y} \frac{1}{N_p} \sum_{p=1}^{N_p} |\mathcal{F}(\hat{\mathbf{y}}_i(n_p))|^2 \quad (7)$$

Here, the set  $\{\hat{\mathbf{y}}(n_p)\}_{p=1}^{N_p}$  denotes ‘‘collocation points’’ for  $\mathcal{F}$ , which are defined as a prediction horizon of  $N_p$  datapoints obtained from the ESN covering the time period  $(T + \Delta t) \leq t \leq (T + N_p \Delta t)$ . Compared to the conventional approach where the regularization of  $\mathbf{W}_{out}$  is based on avoiding extreme values of  $\mathbf{W}_{out}$ , our proposed method regularizes  $\mathbf{W}_{out}$  by using our prior physical knowledge. Equation (7), which is a key equation, shows how to constrain our prior physical knowledge in the loss function. Therefore, this procedure ensures that the ESN becomes predictive because of data training and the ensuing prediction is consistent with the physics. It is motivated by the fact that in many complex physical systems, the cost of data acquisition is prohibitive and thus, there are many instances where only a small amount of data is available for the training of neural networks. In this context, most existing machine learning approaches lack robustness: Our approach better leverages on the information content of the data that the machine learning algorithm uses. Our physics-informed framework is straightforward to implement because it only requires the evaluation of the residual, but it does not require the computation of the exact solution.

### 3 Results

The approach described in section 2 is applied for forecasting the chaotic evolution of the chaotic Lorenz system, which is described by the following equations [7]:

$$\frac{du_1}{dt} = \sigma(u_2 - u_1), \quad \frac{du_2}{dt} = u_1(\rho - u_3) - u_2, \quad \frac{du_3}{dt} = u_1 u_2 - \beta u_3 \quad (8)$$

where  $\rho = 28$ ,  $\sigma = 10$  and  $\beta = 8/3$ . These are the standard values of the Lorenz system that spawn a chaotic solution [7]. The size of the training dataset is  $N_t = 1000$  and the timestep between two time instants is  $\Delta t = 0.01$ .

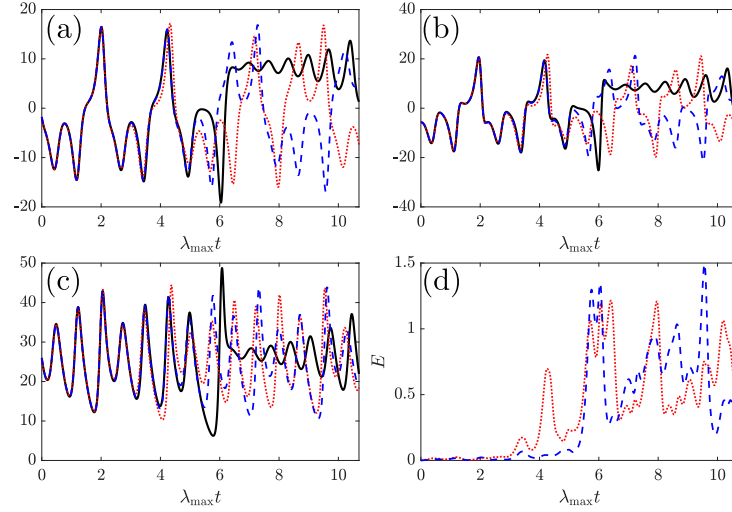
The parameters of the reservoir both for the conventional and physics-informed ESNs are taken to be:  $\sigma_{in} = 0.15$ ,  $\Lambda = 0.4$  and  $\langle d \rangle = 3$ . In the case of the conventional ESN, the value of  $\gamma = 0.0001$  is used for the Tikhonov regularization. These values of the hyperparameters are taken from previous studies [10, 11].

For the physics-informed ESN, a prediction horizon of  $N_p = 1000$  points is used and the physical error is estimated by discretizing Eq. (8) using an explicit Euler time-integration scheme. The choice of  $N_p = 1000$  gives equal importance to the error based on the data and the error based on the physical constraints. The optimization of  $\mathbf{W}_{out}$  is performed using the L-BFGS-B algorithm with the  $\mathbf{W}_{out}$  obtained by Ridge regression (Eq. (4)) as the initial guess.

The predictions for the Lorenz system by conventional and physics-informed ESNs are compared with the actual evolution in Fig. 2, where the time is normalized by the largest Lyapunov exponent,  $\lambda_{\max} = 0.934$ , and the reservoir has 200 units. Figure 2d shows the evolution of the normalized error, which is defined as

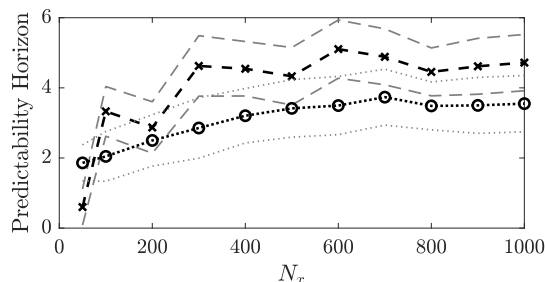
$$E(n) = \frac{\|\mathbf{u}(n) - \hat{\mathbf{u}}(n)\|}{\langle \|\mathbf{u}\|^2 \rangle^{1/2}} \quad (9)$$

where  $\langle \cdot \rangle$  denotes the time average. The physics-informed ESN shows a remarkable improvement of the time over which the predictions are accurate. Indeed, the time for the normalized error to exceed 0.2, which is the threshold used here to define the predictability horizon, improves from 4 Lyapunov times to approximately 5.5 with the physics-informed ESN. The dependence of the pre-



**Fig. 2.** Prediction of the Lorenz system (a)  $u_1$ , (b)  $u_2$ , (c)  $u_3$  and (d)  $E$  using the conventional ESN (dotted red lines) and the physics-informed ESN (dashed blue lines). The actual evolution of the Lorenz system is shown using full black lines.

dictability horizon on the reservoir size is estimated as follows (Fig. 3). First, the trained physics-informed and conventional ESNs are run for an ensemble of 100 different initial conditions. Second, for each run, the predictability horizon is calculated. Third, the mean and standard deviation of the predictability horizon are computed from the ensemble. It is observed that the physics-informed approach provides a marked improvement of the predictability horizon over conventional ESNs and, most significantly, for reservoirs of intermediate sizes. The only exception is for the smallest reservoir ( $N_x = 50$ ). In principle, it may be conjectured that a conventional ESN may have a similar performance to that of a physics-informed ESN by ad-hoc optimization of the hyperparameters. However, no efficient methods are available (to date) for hyperparameter optimization [9]. The approach proposed here allows us to improve the performance of the ESN (optimizing  $\mathbf{W}_{out}$ ) by adding a constraint on the physics, i.e., the governing equations, and not by ad-hoc tuning of the hyperparameters. This suggests that the physics-informed approach is more robust than the conventional approach.



**Fig. 3.** Mean predictability horizon of the conventional ESN (dotted line with circles) and physics-informed ESN (dashed line with crosses) as a function of the reservoir size ( $N_x$ ). The associated gray lines indicate the standard deviation from the mean.

## 4 Conclusions and future directions

We propose an approach for training echo state networks (ESNs), which constrains the knowledge of the physical equations that govern a dynamical system. This physics-informed approach is shown to be more robust than purely data-trained ESNs: The predictability horizon is markedly increased without requiring additional training data. In ongoing work, (i) the impact that the number of collocation points has on the accuracy of the ESNs is thoroughly assessed; (ii) the reason why the predictability horizon saturates as the reservoirs become larger is investigated; and (iii) the physics-informed ESNs are applied to high dimensional fluid dynamics systems. Importantly, the physics-informed framework we propose will be exploited to minimize the data required for training. This work

opens up new possibilities for the time-accurate prediction of the dynamics of chaotic systems by constraining the underlying physical laws.

## References

1. Duraisamy, K., Iaccarino, G., Xiao, H.: Turbulence Modeling in the Age of Data. *Annu. Rev. Fluid Mech.* **51**, 357–377 (2019)
2. Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A.r., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T., Kingsbury, B.: Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Process. Mag.* **29**(6), 82–97 (2012)
3. Jaeger, H., Haas, H.: Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication. *Science* **304**(5667), 78–80 (2004)
4. Jaensch, S., Polifke, W.: Uncertainty encountered when modelling self-excited thermoacoustic oscillations with artificial neural networks. *Int. J. Spray Combust. Dyn.* **9**(4), 367–379 (2017)
5. Krizhevsky, A., Sutskever, I., E. Hinton, G.: ImageNet Classification with Deep Convolutional Neural Networks. *Neural Inf. Process. Syst.* **25** (2012)
6. Ling, J., Kurzawski, A., Templeton, J.: Reynolds averaged turbulence modelling using deep neural networks with embedded invariance. *J. Fluid Mech.* **807**, 155–166 (2016)
7. Lorenz, E.N.: Deterministic Nonperiodic Flow. *J. Atmos. Sci.* **20**(2), 130–141 (1963)
8. Lukoševičius, M.: A Practical Guide to Applying Echo State Networks. In: Montavon, G., Orr, G.B., Muller, K.R. (eds.) *Neural Networks: Tricks of the Trade*. Springer (2012)
9. Lukoševičius, M., Jaeger, H.: Reservoir computing approaches to recurrent neural network training. *Comput. Sci. Rev.* **3**(3), 127–149 (2009)
10. Pathak, J., Hunt, B., Girvan, M., Lu, Z., Ott, E.: Model-Free Prediction of Large Spatiotemporally Chaotic Systems from Data: A Reservoir Computing Approach. *Phys. Rev. Lett.* **120**(2), 24102 (2018)
11. Pathak, J., Wikner, A., Fussell, R., Chandra, S., Hunt, B.R., Girvan, M., Ott, E.: Hybrid forecasting of chaotic processes: Using machine learning in conjunction with a knowledge-based model. *Chaos* **28**(4) (2018)
12. Raissi, M., Perdikaris, P., Karniadakis, G.: Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations. *J. Comput. Phys.* **378**, 686–707 (2019)
13. Raissi, M., Wang, Z., Triantafyllou, M.S., Karniadakis, G.: Deep learning of vortex-induced vibrations. *J. Fluid Mech.* **861**, 119–137 (2019)
14. Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., Hassabis, D.: Mastering the game of Go with deep neural networks and tree search. *Nature* **529**(7587), 484–489 (2016)
15. Wu, J.L., Xiao, H., Paterson, E.: Physics-Informed Machine Learning Approach for Augmenting Turbulence Models: A Comprehensive Framework. *Phys. Rev. Fluids* p. 074602 (2018)